The Mental States of Persons and their Brains

TIM CRANE

Abstract
Cognitive neuroscientists frequently talk about the brain representing the world. Some philosophers claim that this is a confusion. This paper argues that there is no confusion, and outlines one thing that 'the brain represents the world' might mean, using the notion of a model derived from the philosophy of science. This description is then extended to make apply to propositional attitude attributions. A number of problems about propositional attitude attributions can be solved or dissolved by treating propositional attitudes as models.

1. Does the Brain Think?

Consider a picture of a domino with an arrangement of apparently concave and convex circles. The same picture rotated through 180 degrees makes the concavity and the convexity appear reversed. Why does this happen? Why does the very same picture, the same arrangement of pixels or ink on a page, appear so different when turned upside down? Chris Frith gives the following answer in *Making Up The Mind*:

> The light of the sun comes from above … this means that concave objects will be dark at the top and light at the bottom, while convex objects will be light at the top and dark at the bottom. Our brain has a simple rule built into its wiring. It uses this rule to decide whether an object is concave or convex.[1]

Frith claims that the brain has a rule built into it and uses this rule to makes decisions about how things seem. Taken literally, saying that the brain 'uses' a rule, as opposed to merely behaving in a rule-governed or law-like way, implies that the brain somehow represents the content of the rule. And saying that the brain makes decisions implies that the brain is something like a thinker; in short, the brain thinks.

[1] Chris Frith, *Making Up the Mind* (Oxford: Wiley-Blackwell 2007), 128.

doi:10.1017/S1358246115000053 © The Royal Institute of Philosophy and the contributors 2015

For those like Frith, the idea that the brain represents the world (or 'thinks'), should be accepted as part of the orthodox ideology of cognitive science or cognitive neuroscience. Others say that the question is totally confused.[2] For them, this kind of talk embodies a mistake; or even worse, a fallacy (and not all mistakes are fallacies). M.K. Bennett and P.M.S. Hacker have argued that it is an instance of what they call the 'mereological fallacy': the 'mistake of ascribing to the constituent parts of an animal attributes that logically apply only to the whole animal'.[3] In this they take themselves to be following Wittgenstein, who famously said that 'only of a human being and what resembles (behaves like) a living human being can one say: it has sensation; it sees, is blind; hears, is deaf; is conscious or unconscious'.[4] A brain doesn't resemble a living human being; it doesn't even resemble something that resembles a living human being. A chimpanzee at least resembles something that has thoughts; the brain does not even resemble a chimpanzee.

Bennett and Hacker think it's not empirically or straightforwardly false that the brain represents the world. Rather, it is a conceptual truth 'that perception, thoughts and feelings are attributes of human beings, not of their parts – in particular not of their brains'.[5] So it is a fallacy to say something which is incompatible with this conceptual truth. But the supposed fallacy cannot derive from any conceptual principle that you cannot, in general, attribute things to the parts of a system that you would also attribute to the whole. There are many cases where you can do this (e.g. weight, colour etc.) which of course Bennett and Hacker will not deny. So if there is a fallacy here, it must be to do with the use of the terms 'thought' or 'sensation' or 'consciousness' or 'thinking' or 'deciding': mental terms in general.

It is true that the paradigm applications of the concepts of thought, decision, sensation and so on are to organisms: things like human beings and those animals which it makes sense to describe as conscious or thinking. But often we extend the use of words creatively beyond their paradigm applications, to illuminate or illustrate some significant feature of the thing described. Stephen Mulhall makes this point in a recent discussion of the concept of a picture.

[2] See M.R. Bennett and P.M.S. Hacker, *Philosophical Foundations of Neuroscience* (Oxford: Wiley-Blackwell, 2003).

[3] Ibid., 72.

[4] Ludwig Wittgenstein, *Philosophical Investigations* (Oxford: Blackwell, 1953), §281.

[5] M.K. Bennett and P.M.S. Hacker, *Philosophical Foundations of Neuroscience*, 3.

In considering what he calls the 'projectability of language', Mulhall writes,

> The word 'picture' denotes, among other things, abstract paintings, representational paintings and films, i.e. motion pictures. Although abstract paintings and films each have something in common with representational paintings, there seem to be no relevant features common to a Jackson Pollock drip painting and a projected image of Humphrey Bogart, and yet we have no inclination – do we? – to say that the word has one meaning in a conversation about *Casablanca* and another when the talk turns to *Lavender Mist*.

He continues,

> If someone were to construct such a pattern of use from scratch, we might find it rather puzzling. But if we see it as a process of historical development, the puzzle dissolves. In the case of 'picture', the original focus on representational painting naturally licenses an extension of the term's use to photographs and thence to motion pictures; and developments in painting also made natural a different extension of the term to include canvasses of a non-representational sort.[6]

As the use of the word develops across time, we can extend its use by applying it intelligibly to other things. We find it very natural to extend the use of a word beyond some original or initial contexts, without the word changing its meaning in any strict sense. Another simple example is the concept of flying. Suppose a child asks, can a person fly? Of course people can't fly, we might reply; Superman can't really fly, it's only a story. So what is it to fly? Perhaps we might say something like this: to be propelled through the air, by using the motion of wings (as a bird flies) or some other kind of mechanism attached to the flying object (as a rocket flies).

But what if I tell you that I am flying to Turin in a few weeks' time? Then we should give another answer to the question 'can people fly?'. Of course people can fly: travelling in an aeroplane is flying. It would be at best a bad joke to respond to the question 'did you fly here, or did you come by train?' by saying 'no, I can't have flown here, because to fly is to propel yourself through the air using wings or some other kind of mechanism'. And yet, as Mulhall says about

[6] Stephen Mulhall, Stanton lectures 2014, University of Cambridge (unpublished).

pictures, there should be no temptation to say that the word has one meaning in the first conversation and a different meaning in the second. Just as developments in painting led to the natural extension of the word 'picture' to Pollock's work, so developments in aeronautical technology led to a natural extension of the word 'fly' to what people do when they go somewhere by plane: in this sense, people can fly. When we focus on the historical development of the use of the word, any puzzlement we felt about giving different answers in the two scenarios should vanish.

As Mulhall says, this possibility of words having uses which extend across many different kinds of case, without being ambiguous or polysemous, is something to which Wittgenstein drew attention:

> I can think of no better expression to characterise these similarities than 'family resemblances'; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, and so on and so forth – overlap and criss-cross in the same way. – And I shall say: 'games' form a family.[7]

The relationship between the different uses or meanings of the words 'picture' and 'fly' in our scenarios seems to be a kind of family resemblance of the sort Wittgenstein talks about. To say this is not to propose or defend any particular theory or account of meaning; it is simply to draw attention to a phenomenon. But once we have recognised this phenomenon, we thereby open up the possibility of making sense of the idea that the brain thinks or represents the world.

However, some may object at this point that the meaninglessness of saying that the brain thinks is underwritten by the well-established distinction between the personal and sub-personal levels of psychological explanation and description. The concepts of thought, decision, and so on belong to the personal level, and we should not extend them to the sub-personal level. Before explaining what I think it means to say that the brain thinks, I need to put this objection to one side.

Daniel Dennett introduced the personal/sub-personal distinction in *Content and Consciousness* in the context of a discussion of pain:

> When we've said that a person has a sensation of pain, that he locates it and is prompted to react in a certain way, we have said all there is to say within the scope of this vocabulary. Since the introduction of unanalysable mental qualities leads to a premature end to explanation, we may decide that such an

[7] Wittgenstein, *Philosophical Investigations*, §67.

> introduction is wrong, and look for alternative modes of explanation. If we do this, we must abandon the explanatory level of people and their sensations and activities, and turn to the sub-personal level of brains and events in the nervous system.[8]

As Dennett's remark about 'alternative modes of explanation' indicates, the personal and sub-personal are two modes of explanation of the one cognitive system. The personal mode of explanation appeals to concepts like *sensation* and then *behaviour*, *location* and *reaction*; and the alternative explanatory level of the sub-personal appeals to brains, nerves, neurones, synapses and events in the nervous system.

Dennett points out that the lesson about the personal/sub-personal distinction has occasionally been misconstrued 'as the lesson that the personal level of explanation is the *only* level of explanation when the subject matter is human minds and actions'.[9] Now a strict and literal reading of Wittgenstein's remark about 'only of a human being and what resembles a living human being…' would presumably entail this claim that the personal level is the only level of explanation that matters when the subject matter is human minds and actions. But this is not Dennett's view. In his commentary on Bennett and Hacker's idea of a mereological fallacy, he writes,

> we don't attribute fully-fledged beliefs to the brain parts. That would be a fallacy but we attribute an attenuated sort of belief to these parts, stripped of many of its everyday connotations. Just as a young child can sort of believe that her daddy is a doctor, without full comprehension of what a daddy or a doctor is, so a robot, or some part of a person's brain can sort of believe that there is an open door a few feet ahead… far from being a mistake to attribute hemi-semi-demi-proto-quasi-pseudo-intentionality to the mereological parts of persons, it is precisely the enabling move that lets us see how on earth to get the whole wonderful persons out of brute mechanical parts.[10]

Dennett's view is clearly that the distinction between the personal and the sub-personal levels of explanation is clearly compatible with two ideas: (i) that we might extend to the sub-personal level

[8] Dennett, *Content and Consciousness* (London: Routledge and Kegan Paul 1969) 95.

[9] Ibid., 95.

[10] Daniel C. Dennett, 'Philosophy as Naive Anthropology: Comment on Bennett and Hacker' in *Neuroscience and Philosophy* (New York: Columbia University Press, 2007), 87–9.

words whose paradigmatic application is at the personal level; (ii) that we might use this extension to help us explain how intentionality at the personal level is possible. One could reject either or both of these ideas, but this rejection is not implied by the very distinction between the personal and the sub-personal.

So this brings us back to our question, what does it mean to say that the brain thinks, or represents the world? I take Frith's comments quoted above as representative: many cognitive neuroscientists and others are perfectly happy to talk in terms of the brain representing the world. Given this, Hacker's approach leaves us with a mystery: how is it that so many people, apparently highly competent in their use of language and highly knowledgable about the empirical facts, fall so easily into such a simple fallacy? What is more, why is it that they are so resistant to recognising that they have made this apparently simple mistake?

I will argue in the rest of this paper that in fact it is not a mistake, and I will attempt to explain what it means to say that the brain represents the world. Just as it makes perfect sense to say that people fly, and it makes perfect sense to say that a Jackson Pollock painting is a picture, so similarly it makes sense to say that the brain represents the world, or even that it thinks. If we allow for words to extend their meaning through a historical process, then we can see how this is not obviously meaningless. In fact, on broadly Wittgensteinian grounds, we should accept that such transfers and extensions of meaning are part of the essence of our language, and there is nothing in principle stopping the talk of neuroscientists being such a case. But how should we spell out this idea that the brain represents?

2. Contents as Models

One answer to this question, going back to the 1970s, is that representation in the brain involves there literally being symbols written in your brain: there are symbols which represent the contents of your beliefs and other personal mental states, or your sub-personal states, or both. These symbols are part of what's known as a 'language of thought'; this view was defended by Jerry Fodor in 1975.[11] Fodor

[11] Jerry A. Fodor, *The Language of Thought* (Hassocks: Harvester 1975). For a critical overview, see Susan Schneider, *The Language of Thought* (Cambridge, Mass.: MIT Press 2011).

argued that there can't really be beliefs unless there are representations in the brain.

The main problem with the Language of Thought hypothesis, to be blunt, is that there is no reason to believe there is such a thing as the Language of Thought. Some people hypothesise that there might be such a thing for the sake of argument, but few of those who say this never defend the idea that it really exists (and even Fodor himself is sceptical these days[12]). As Frances Egan remarks: 'There isn't much empirical support for this view. It's a very elegant picture but it's not likely to be the way that minds developed naturally, as a product of evolution, in fits and starts'.[13]

Instead of speculating about the inner structure of the brain, or simply denying that representation in the brain makes any sense, we should ask instead of what it is that people are doing when they attribute representations to the brain – when they say things like 'the brain knows that light comes from above'. What is it that those who attribute intentional states are actually doing in this kind of case?

The interpreter is the theorist – the neuroscientist or psychologist – who is trying to provide the best explanatory structure to account for how the system moves from one state to the next. The system, in the cases we are considering, is the brain, and what the theorist is trying to do is to explain why the system moves from one state to another. There are states of the system which the theorist can then map on to what I'm going to call its contents. The content might be something like the rule that *if the object is light at the top and dark at the bottom then this is a convex object*, for example. Contents are related to the intrinsic or non-intrinsic states of the system by a *mapping* provided by the theorist; that is to say, a correlation between the states of the system and the content. I suggest that we think of the content as part of a *model* of the system, in the sense in which this word is used in the philosophy of science.

It is a familiar claim in recent philosophy of science that scientific theories are 'collections of models'; but given the variety of things the word 'model' has come to mean, some clarifications are needed. The original proposal that theories should be thought of in terms of models is normally traced back to Patrick Suppes's work in the 1960s, subsequently developed by Bas van Fraassen.[14] Suppes and

[12] Jerry A. Fodor, *LOT2* (Oxford: Oxford University Press, 2008).

[13] Frances Egan, http://www.3ammagazine.com/3am/meaning-as-gloss/.

[14] Patrick Suppes, 'A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences' *Synthese* **12** (1960), 287–301;

van Fraassen used the word 'model' in the sense of model theoretic semantics; a model of a theory is a collection of objects which renders true the claims of the theory. Hence the association of the idea of a theory as a collection of models with the label 'the semantic view of theories', contrasted with the 'syntactic' view of theories defended by the logical empiricists. 'Semantic' is appropriate because scientific models were being conceived of in terms of model theory, the standard semantic framework for formal languages.

Later work on models by Ronald Giere and others emphasised something quite different: the use of simplified mathematical structures (e.g. equations describing ideal populations), imagined comparisons (the atom is like a solar system) or even concrete objects (the actual wire and wood construction which represents the double-helix structure of DNA).[15] All these things are classified as models by scientists, and philosophers of science attempted to make sense of them. But models in this second sense look very different from the models of model theory, as a number of writers have emphasised.[16]

In model theory, a model is collection of objects and operations on these objects – a set-theoretic structure – which makes the sentences of a theory true. (Nothing metaphysically weighty is meant by 'making true' here; this is just a standard definition of a model.) Even if model theory can be used to illuminate scientific theorising, as Suppes and van Fraassen argue, it is plain that a model-theoretic model does not look much like Rutherford's solar system model of the atom. For the solar system does not make true any claims about atoms; and nor does the comparison between the solar system and the atom make this true. Rather, the solar system is used to represent an aspect of how things are with the atom, just as equations describing an idealised population in biology is used to represent an aspect of how things are with some real population. Neither the equations

---

Bas Van Fraassen, *The Scientific Image* (Oxford Oxford University Press, 1980).

[15] See e.g., Ronald N. Giere, 'Using Models to Represent Reality' in *Model-Based Reasoning in Scientific Discovery*, (ed.) L. Magnani, N. J. Nersessian, and P. Thagard (New York: Kluwer/Plenum, 1999), 41–57.

[16] See Stephen Downes, 'The Importance of Models in Theorizing: a Deflationary Semantic View' in Hull D, Forbes M, Okruhlik K (eds) *PSA* 1992, vol. 1. Philosophy of Science Association, East Lansing (1992), 142–153; and Martin Thomson-Jones 'Models and the Semantic View' *Philosophy of Science* **73** (2006), 524–535.

nor the idealised population can be described as making true the claims about the real population.

The model-theoretic conception of models makes more sense if we think of the model as the *comparison* between the solar system and the atom. Here we might consider the model to be the mapping itself (sometimes described as an isomorphism, as it is by Suppes) between aspects of the real-world system and those of the object used to represent it. Whether this is a good way to understand models in science is much discussed in the philosophy of science; but what is clear is that it is a different idea from that of those who talk of one system being a model of another. Peter Godfrey-Smith puts this point well:

> Representation of a real-world system involves two distinct relations, the specification of a model system and some relevant similarity between model system and the world itself. So the word 'model' tends to be ambiguous here, between what we can call the model system and the model description.[17]

In what follows, I will talk of models as being what Godfrey-Smith calls the 'model system', and not as the 'model description' (i.e. the description mapping the system to the real-world system under investigation). To the extent, then, that the semantic conception of theories is tied up with the model-theoretic conception of models, we should avoid talking about 'the semantic conception' of theories here.

Models, in the sense I intend, are used to understand the behaviour of real-world systems by being used to represent how things are with those systems at a time, or how they evolve across time. Typically, the models are simpler than the real-world systems under investigation. They may involve idealisations (frictionless planes), or even empirical falsehoods (rational actor models in economics), and they may be unspecific in certain respects (a model of a cell may leave out information relating to what kind of cell it is).[18] The point of the model is to facilitate understanding of the real-world system by examining the behaviour of the model system: it is what Michael Weisberg calls an 'indirect theoretical investigation of a real-world phenomenon'.[19]

[17] Peter Godfrey-Smith 'The Strategy of Model-Based Science' *Biology and Philosophy* **21** (2006), 725–740; 733.

[18] See Downes, 'The Importance of Models in Theorizing: a Deflationary Semantic View', 145–6.

[19] 'Who is a Modeler?' *British Journal for the Philosophy of Science* **58** (2007), 207–233; 208.

If we apply this to the case of the brain, the following picture emerges. The brain is the system under investigation (the 'real-world system') and the theorist attributes to it certain states, in order to predict or explain certain outputs. The theorist models these states by relating them to the abstract objects which are what I call the contents of the system. They expect to give a better understanding of the transitions between states of the brain by relating these states to contents than they would by merely citing neurochemical interactions, say, or gross external behavioural changes. This is what is going on when Frith claims that 'the brain has this simple rule built into its wiring; it uses this rule to decide whether an object is concave or convex'. The rule is that because light generally comes from above, concave objects will be dark at the top and light at the bottom; convex objects will be light at the top and dark at the bottom. That doesn't mean that those words are written in the brain, that those words are written in English or any other language, or even in the Language of Thought. The attribution of content is, rather, an abstraction away from the activity of the brain in a way that can help you with predicting and explaining what's going on. The claim, then, is that the process in the brain is understood by modelling it with a rule relating shading to convexity and concavity.

As Giere has emphasised, modelling works by exploiting relations of similarity.[20] What is similar to what, in the case of the brain? It's not that the brain state, whatever it is, resembles an abstract object, any more than a population resembles the equations used to model its growth. It's rather that the movement between states is similar to the stages of an explicit inference from the rule about concrete objects, plus the input from the image, to the conclusion about how the object looks. Inferences relate propositions; so the claim is that what is going on in the brain resembles a relationship between propositions. The appropriate comparison here is between modelling the brain with an inference, and modelling some target system by using a mathematical model.

Notice that this idea contains an echo of Fodor's famous argument for the Language of Thought based on the nature of mental processes.[21] I agree with Fodor that there is this similarity; but I resist the jump to the conclusion that this gives us a reason to say that there are symbols in the brain. Once equipped with a proper

[20] See Giere, 'Using Models to Represent Reality'.

[21] See Tim Crane, *The Mechanical Mind* (London: Routledge, 2003) chapter 4, for an exposition of this argument.

understanding of models, there is no need to move to this more outlandish hypothesis.

This does however raise an important question. What is the difference between a theory which models the transitions among states of the brain by relating them to representational contents (i.e. treats the states as representations) and one that merely treats the transitions as law-governed processes? When looking at the Kanizsa triangle, for example, normal perceivers see a white triangle as occluding segments of three black circles at its corners. Some neuroscientists and psychologists talk about the representation being 'completed' in the visual system by means of it making certain 'assumptions' about objects and how they normally relate to one another. On this understanding, the brain is conceived of as making an inference: in moving from state to state, its states are modelled by contents. But on another understanding, the brain is simply a law-governed system which produces certain visual outputs by being governed by a law about how objects normally look. On this view, there is no reason to call this an inference, not least because an inference must be rationally sensitive to further information, and this one isn't: no matter what you know, what you see will not change.

To determine whether it is correct to model a brain process by relating it to an inference will depend on the details of the case. My aim here is not to defend any particular hypothesis that treats the brain as making an inference; my aim is only to argue that it makes sense, and to say something about what kind of sense it makes.

3. The Propositional Attitudes

My discussion so far has been about things going on the sub-personal level; but does the modelling picture apply to the states of the whole person? In particular, does it apply to beliefs, desires, hopes and the other propositional attitudes? Some philosophers may agree that the modelling picture makes sense applied to sub-personal states but not to personal-level states; perhaps because these states have 'original' rather than 'derived' intentionality.[22]

Here I want to resist this objection. In the remainder of this paper, I will argue that ascriptions of propositional attitudes employ propositions as models in a similar way that ascriptions of sub-personal states employ contents as models. But I don't say this because I

[22] For this distinction, see John R. Searle, *Intentionality* (Cambridge: Cambridge University Press, 1983).

reject the distinction between original and derived intentionality. Rather, I would draw the significant distinction in this area not between sub-personal states and personal-level propositional attitudes, but between intrinsically intentional states of consciousness and all the other mental states. However, I cannot defend this further thesis in this paper; so I will confine myself to the claim about models.[23]

Contemporary philosophy of mind contains two influential views about the propositional attitudes. The first is that the concept of intentionality, the mind's representation of the world or direction upon its objects, should be understood entirely in terms of the propositional attitudes. This is a thesis that's been held by many people – Donald Davidson and many others. I call it 'propositionalism'.[24]

The second view – which I will call the 'relational thesis' – is that the propositional attitudes should be thought of as (literally) relations to propositions: a propositional attitude like *believing that the sun is shining* is a relation between the person and the thing believed, namely, the proposition that the sun is shining. When you believe that the sun is shining you have one kind of relation to it, and when you hope that the sun is shining you stand in another kind of relation to it. The conjunction of propositionalism with the relational thesis implies that intentionality should be understood in terms of relations to propositions.

For a long time the relational thesis has been something of a dogma in analytical philosophy of mind. Fodor made his argument for this thesis a cornerstone of his intentional realism: 'Believes looks like a two place relation so it relates two things the thinker and the proposition and it would be nice if our theory of belief permitted us to save the appearances'.[25] And in a famous paper, Hartry Field said a similar thing: 'propositional attitude attributions appear to relate people to non-linguistic entities called propositions' and he then claims that this fact is a problem for materialists.[26] The picture we are given is of a real relation to a proposition – the proposition is the thing you believe, and your belief state is a relation to it.

[23] For defences of the latter thesis, see John Searle, *The Rediscovery of the Mind*, and Galen Strawson, *Mental Reality*.

[24] See Donald Davidson, 'Mental Events' in *Essays on Actions and Events* (Oxford: Oxford University Press, 1982). In chapter 4 of *The Objects of Thought* (Oxford: Oxford University Press 2013) I offer a critique of propositionalism which is independent of the present paper.

[25] Jerry A. Fodor, 'Propositional Attitudes' *The Monist* **61** (1978) 501–23.

[26] Hartry Field, 'Mental Representation' *Erkenntnis* **13** (1978) 9–61, 10.

Before deciding whether this claim is true, there is a prior question: what does it mean? What does it mean to say that you're 'related to a proposition by the relation of believing', if that is not just another way of saying that you believe it? What's the point of saying that you stand in a 'real relation' to a proposition?

The picture I want to reject is that it is a basic psychological or metaphysical fact, something that needs to be explained, that people are related to propositions. But this does not mean that it is not true. Rather than being something that needs to be explained, it is rather part of the theoretical explanation or description of your state of mind. In other words, we should think of the ascription of propositional content to a person – placing someone in a relation to a proposition – as a way of modelling their mental state.

When we say that someone believes that *p*, or thinks that *p*, what we are doing is picking out a feature of their state of mind – their whole psychological outlook, or world picture – by relating it to this abstract object, the proposition *p*. The proposition serves to pick out part of the way the subject represents the world. We use a sentence to express the proposition, but we shouldn't think of the belief simply as a relation to a sentence, since the same feature of the state of mind can be expressed by different sentences (in the same or different languages). So for example, if we say that someone thinks Scotland should be an independent country, we are picking out an aspect of their world view, by relating them to the proposition expressed by the sentence 'Scotland should be an independent country' and all sentences that mean the same.

It is not plausible that someone could believe this without believing, for example, that some countries should be independent; this is an instance of the well-known 'holism of the intentional', which should be accepted by everyone. This means that if we attribute the first belief we should also attribute the second, unless we have some countervailing reason not to. Why this is so, is a question which divides theories of mind and intentionality: 'inferentialists' take these facts to be the *basis* of intentionality, 'representationalists' take them to be *explained* by facts about intentionality. Here my aim is not so much to contribute to this debate but to point to the role of 'the proposition' in describing beliefs and the relationships between them.

In identifying a belief by using a proposition, we are exploiting the logical and conceptual relationships between propositions to reflect the holism of the intentional. The fact that if you are ascribed the belief that Scotland should be an independent nation means that you can also be ascribed the belief that some countries should be

independent mirrors the fact that the proposition that *Scotland should be an independent nation* entails that *some nations should be independent*. I presuppose here that propositions stand in logical and conceptual relationships – they are inconsistent, they entail one another, they support one another and so on. On the assumption that we have a fairly good understanding (or at least an agreement) about which logical and other relationships hold between propositions, we can then use this understanding or agreement to underpin our conjectures about propositional attitudes. It is because we understand that *A* entails *B* that we can use this understanding to say that 'if someone believes that *A* then they will/should believe that *B*'.

This is why I say that the relation to a proposition is a way of modelling a belief. Just as Rutherford used our antecedent understanding of the structure of the solar system to model or picture something we are trying to understand – the atom – so we can use our antecedent knowledge of the relations between propositions to model the relationships between beliefs, and thus various aspects of the beliefs themselves. This is one feature of models: they use something which is in some way already understood to enlighten us about something that is less understood. Models do this by idealisation. If you say that someone believes that Scotland should be an independent nation, you do not thereby say everything about their conception of Scotland, nations and independence. They may have a very complex conception which is not easily summarised by a single sentence; yet it may still be perfectly true that they believe this, and in that sense the proposition models what they believe. As Robert Cummins puts it, 'attributing a belief is going to be a bit like attributing a point of view to an editorial'.[27] You can attribute a point of view to an editorial even if in doing so, you don't actually pick out a sentence that is contained in that editorial.

I mentioned above that when we attribute a belief, we typically commit ourselves to attributing certain related beliefs; or rather, we commit the believer to having certain related beliefs. This is the holism of the intentional. But of course, people do not always believe the logical consequences of what they believe, and our belief ascriptions should reflect this fact. Thinking of propositions as modelling beliefs captures this situation: the model is an idealisation. We say that if someone believes *A* then they should believe *B* but that may be because proposition *A* entails proposition *B*, or because it gives strong inductive support to *B*, or because it provides some other

[27] Robert Cummins, *Meaning and Mental Representation* (Cambridge, MA: MIT Press, 1989), 144.

kind of reason for believing *B*. We might expect that a thinker will have the second belief if they have the first; this is a reasonable idealisation. But there might be all sorts of reasons why they don't have the second, and if so we would have to revise our model in the light of other things we find out about them. (I do not mean to imply here that thinkers are obliged to believe all logical consequences of what they believe, only that there are some cases where they do have an obligation to believe some of the consequences.)

There are two central questions about the propositional attitudes which this modelling approach answers. The first concerns what kind of objects propositions themselves are; the second concerns which specific propositions (however conceived) actually are the contents of any particular belief.

The first question is about the metaphysics of propositions. Some philosophers (Russellians) think that the constituents of propositions are objects and their properties; some (Fregeans) say that the constituents of propositions are senses or 'modes of presentation'; others (Lewisians/Stalnakerians) say that a proposition is a set of possible worlds.[28] Which view is right? Which objects are thinkers related to in their propositional attitudes?

There are things to be said in favour of each view. For example, sets of possible worlds are formally tractable – relations of entailment (etc.) can be understood set-theoretically. But there are also aspects of the possible worlds view of propositions which are very bad for representing distinctions between mental states. As has been observed many times, if a proposition is a set of possible worlds, then a necessarily true proposition is the set of all possible worlds. So all necessarily true sentences express the same proposition; and all necessary false sentences express the same proposition. We could take this to be a *reductio ad absurdum* of the view, but I don't think we should do this. Rather, I think what it shows is that propositions as sets of possible worlds constitutes a partial model of certain states of mind. It is a model which is good for some things and not for others.

A similar thing can be said about the view that propositions are Russellian. This too is a partial model. It might express the fact that the object of your thought is that *very particular object* you are thinking about, and no other object will do. Similarly the Fregean view that distinguishes propositions (*Gedanken*) more finely than objects and properties may be used to express a finer-grained perspective on a subject's mental life. But we should resist the idea

[28] For a recent contribution to this debate see Jeffrey C. King, *The Nature and Structure of Content* (Oxford: Oxford University Press, 2007).

that only one of these views is correct in giving the content of someone's belief, that *this* and not *that* is the proposition they are related to. The modelling picture allows that people can be related to propositions of all these kinds; in other words, objects of all these kinds can be used to pick out their beliefs.

In a recent insightful discussion of this subject, Ian Rumfitt says:

> It is pretty clear that the maxims we ordinarily go by in reporting the speech and beliefs of others do not place people in relation either to Fregean *Gedanken* or to Russellian propositions; for that reason, there is little mileage in discussing whether *Gedanken* or Russellian propositions best match our pre-theoretic notions of saying or believing the same thing. These entities are better conceived as constructs, postulated for various theoretical purposes in philosophy, linguistics and psychology. The proper topic of debate, then, is whether a given construct serves a specified theoretical purpose. It is entirely possible that Fregean *Gedanken* might best serve one such purpose, Russellian propositions another, and indeed Stalnakerian propositions (i.e. sets of possible words) a third.[29]

Unlike Rumfitt, I am happy to say that ordinary belief and speech reports do place thinkers in relation to propositions, since I think that ascribing a belief is modelling, and this places the thinker in a relation to a proposition, in a metaphysically innocuous sense. But this is a minor disagreement; if we read Rumfitt as saying that our ordinary maxims for reported speech and belief do not *determine* that one kind of proposition rather than another are the objects of belief or speech, then his remarks here get the matter exactly right.

The second question which the modelling picture answers is about which specific proposition actually is the content of a given belief, once we have settled on a general type of proposition. How specific or determinate should the content of a belief be, for example? And how should we identify the concepts which a thinker employs when they have a given belief? These questions arise particularly starkly in connection with animal belief. Consider Norman Malcolm's example of whether a dog believed that a cat went up a certain tree.[30] Some say that if the dog believes that the cat ran up the tree, this requires that the dog has the concept of a tree, and this requires

[29] Ian Rumfitt, 'Truth and Meaning' *Proceedings of the Aristotelian Society, Supplementary Volume* **88** (2014), footnote 6.

[30] Norman Malcolm, 'Thoughtless Brutes' *Proceedings and Addresses of the American Philosophical Association* **46** (1973), 5–20; cf. Donald

that it can think of a tree as something living, and the tree has leaves and the tree is made of wood … Most or all of these things we believe about trees. But how can a dog believe any or all of those things? The same applies to the beliefs of children. When a child believes – to use Dennett's example – that her father is a doctor, what exactly does she believe? If beliefs are determinate states of subjects, then surely there must be an answer to this question.

The modelling view gives an answer. The child and the dog have a conscious perspective on the world, a point of view. What we do when we ascribe them a belief is to attempt to make sense of their behaviour (running towards the tree, saying 'Daddy is a doctor') by identifying some feature of their point of view. This feature we pick out using a proposition. The proposition can truly describe part of that point of view – after all, it is a *tree* that the dog is thinking about, and it is *being a doctor* that the child ascribes to her father. But that does not require that the dog or the child means *tree* or *doctor* in exactly the same sense that we do. The relation to the proposition is a partial model.

And the same applies to more sophisticated thinkers. There is normally not just one sentence that expresses what you believe – in other words, there are many propositions that characterise your belief in different ways. As you struggle to express your beliefs, what you're trying to do is fix on what is the most appropriate way of expressing them. This does not mean that there is no fact of the matter about what you believe, that it is 'simply' a matter of interpretation. I say there is such a fact; but it is often a very complex fact, because of the holism of the intentional. Even where the simplest beliefs about your perceived environment are concerned, what you believe about here depends on so many beliefs about other things, that we should not expect that one sentence can adequately capture it all. It might be replied that a very *long* sentence could do it; maybe, but the suggestion is practically worthless when it comes to understanding real belief ascriptions.

It should be obvious that the view I am defending here owes a lot to Daniel Dennett's views about intentionality and belief-ascription.[31] But this does not mean that the view is some sort of anti-realism or instrumentalism about the mental (regardless of where Dennett himself stands on this vexed issue). On the contrary; I claim that my view involves a robust psychological realism. There is a

---

Davidson, 'Thought and Talk' in *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984).

[31] See in particular, 'Beyond Belief'.

psychological reality out there: this is the reality that you're attempting to model by using propositions in these various different ways. And just as the modelling conception of the atom is compatible with a hard-headed realism about atoms, so the modelling conception of the propositional attitudes is entirely compatible with realism about the psychological. Psychological reality is precisely that which we are trying to model in our personal and sub-personal psychological ascriptions.[32]

University of Cambridge
tc102@cam.ac.uk

[32] Thanks to Ali Boyle, Dan Brigham, Katalin Farkas, Anthony O'Hear and Michael Weisberg for discussion, to members of the audience at the Royal Institute of Philosophy for helpful comments at the RIP meeting in February 2014, and to Stephen Mulhall for permitting me to quote from his unpublished work. An earlier version of this talk was given at the University of London's Institute of Philosophy in June 2012, at a workshop on Dennett's personal/sub-personal distinction; thanks to Dan Dennett for his comments on that occasion.